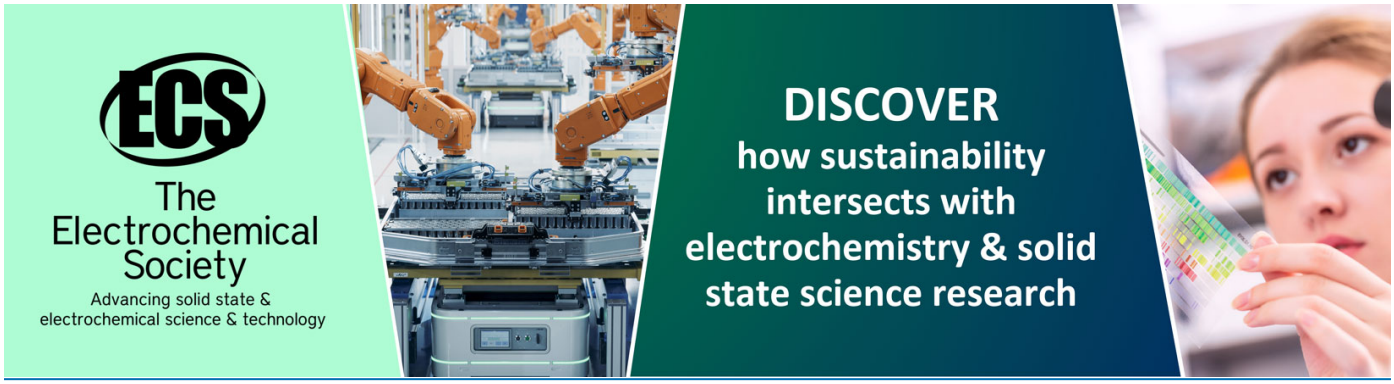
Bridge Damage Detection and Recognition Based on Deep Learning



Bridge Damage Detection and Recognition Based on Deep Learning

Xiuxin Chen, Yang Ye*, Xue Zhang and Chongchong Yu

Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China

*Corresponding author email: 1536849703@qq.com

Abstract. Bridge damage detection is of vital importance to bridge safety. Nowadays the damage detection is mainly performed by human which is inefficient. We pro-posed a bridge damage detection and recognition method based on deep learning which is named DT-YOLOv3 in this paper. Our method is based on YOLOv3 object detection method and several improvements were made. First, deformable convolution was used to extract more accurate features, and transfer learning was introduced to improve the detection accuracy. Then, the model was compressed using group convolution and pruning. The test results show that our method is more effective than state-of-the-art methods and costs less time.

1. Introduction

With the rapid development of transportation system in China, bridges are of great significance in China's transportation system. Bridge damage detection can timely find the appearance damage of bridges and provide valuable information for bridge maintenance to ensure the safety of bridges. Nowadays, the bridge damage detection is mainly performed by workers with their eyes which is inefficient. Deep learning technology has achieved good results in the field of complex image detection and recognition in recent years. In 2012, AlexNet [1] extracted the two-dimensional features of images by constructing Convolutional Neural Networks (CNN). The Regions with Convolutional Neural Network Features (RCNN)[2] proposed in 2014 is the first algorithm applying deep learning in object detection. In 2015, Fast-RCNN [3] re-moved the SVM classification mechanism in RCNN, and in the same year, Faster-RCNN[4] was proposed, which focused on the Region Proposal Network (RPN) method to replace the complex heuristic object Region search. The You Only Look Once (YOLO) network [5] proposed in 2015 adopted different solutions. It adopted the idea of regression to directly predict the location coordinates, confidence and classification results of the detected object. However, the detection and recognition accuracy of the detected object is low. The Single Shot Multi Box Detector (SSD) [6] proposed in 2016 used the idea of anchor boxes to improve the recognition accuracy which is better than YOLO and Faster-RCNN. In 2016, the improved algorithm YOLOv2[7] was proposed. It performs poorly in small object detection, but the over-all detection and recognition accuracy is higher than that of SSD. In 2018, YOLOv3[8] was proposed. It used multi-scale prediction to improve the detection accuracy of small and medium-sized objects and adopted regression instead of softmax for classification which can support multi-label classification of objects and achieve better detection and recognition effect.

Object detection methods mentioned above need a large number of training images to achieve good results, while there are few bridge damage images in practice which makes it a few-shot learning problem. In this paper, we first apply several image processing methods to get more bridge damage

images. Then, YOLOv3 is adopted as the basic model for bridge damage detection. In order to effectively extract the damage characteristics such as cracks and collapses, deformable convolution is used. To solve the problem of insufficient damage images, some parameters obtained through training the model using common data sets are transferred to the bridge damage detection model. Then, we use SGD, NAG, and Adam to optimize the model. At last, group convolution and pruning are used to compress the model.

2. Methodology

2.1. YOLOv3 Model

In recent years, the YOLOv3 model has achieved good results in the field of object detection and recognition. YOLOv3 model has higher recognition accuracy and shorter time.

The structure of the YOLOv3 model is shown in Figure 1. The feature extraction layer is from layer 0 to layer 74 and consists of 1×1 and 3×3 convolution layers and residual modules.

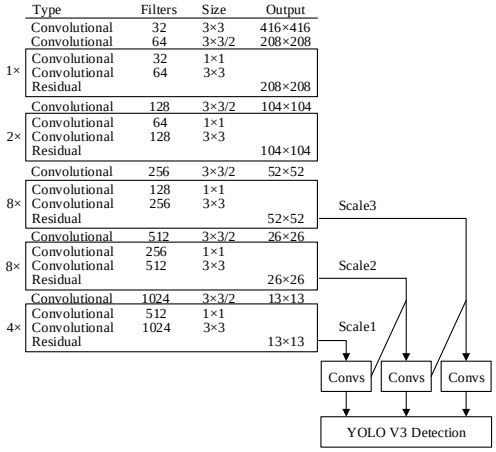


Figure 1. YOLOv3 model structure.

2.2. DT-YOLOv3 Model Combining Deformable Convolution and Transfer Learning

The feature extraction layer of the YOLOv3 network model consists of a 1 × 1, 3 × 3 convolution and residual module [9]. The ordinary convolution mode is difficult to extract complex bridge damage features. Therefore, we add deformable convolution [10] layers after the feature extraction layers of YOLOv3 to better learn the bridge damage features. The upper sampling of high-level semantic features obtained by deformable convolution is fused into feature information of different scales. So in the final features, there are both uniformly sampled features (ordinary convolution) and non-uniformly sampled (deformable convolution) features, which enriches the feature information and facilitates the detection and recognition of the model. As shown in Figure 2, deformable convolution kernel is not limited to the regular point of ordinary convolution, and can randomly sample near the current position.

(a) Ordinary convolution (b) Deformable convolution

Figure 2. Ordinary convolution kernel and Deformable convolution kernel.

There are few bridge damage images, which cannot meet the training requirements of the model. Whil e the image data features have some general probability distributions, especially the features learned at the bottom layers of the deep learning net-work are usually the common features of all images. Theref ore, we save some of the bottom layer parameters which are trained using general image datasets and u se them for the bridge damage detection model to realize knowledge transfer. Then we fine-tune the m odel using bridge damage images. The validity of the transfer method and the number of layers used fo

r transfer learning are verified by experiments. We refer to the YOLOv3 model that combines deforma ble convolution and transfer learning as the DT-YOLOv3 model.

2.3. DT-YOLOv3 Model Compression Based on Group Convolution and Pruning

2.*3.1. Model Compression Based on Group Convolution.* The introduction of deformable convolution layers in the model causes more parameters, so we use group convolution [11] to compress the DT-YOLOv3 model. All feature maps extracted by the final ordinary convolution layer are divided into N groups. Then, each group of feature map is calculated by deformable convolution, and finally the features of all groups are merged.
The model structure after introducing group convolution is shown in Figure 3.

2.*3.2. Model Optimization.* For the model optimization, this paper uses the Newtonian Accelerated Gradient (NAG) algorithm. The algorithm adjusts the parameters of the model using equation (1).

$$\begin{cases} v_t = \gamma v_{t-1} + \eta \nabla J\left(W_t - \gamma v_{t-1}\right) \\ W_{t+1} = W_t - v_t \end{cases} \tag{1}$$

In equation (1), is the decay rate, is the learning rate, $\nabla J\left(W_t - \gamma v_{t-1}\right)$ is the change amount of the gradient at time t and the gradient at time t-1. It can be seen that the update of the model parameters not only depends on the gradient of the current time, but also that of the previous time. This can reduce the influence of noise and is actually a proximation of second-order derivative of the loss function.
For the optimization of the learning rate, we choose the Adam algorithm in this paper. We set the independent adaptive learning rates for different parameters by calculating the first and second moment estimates of the gradient.

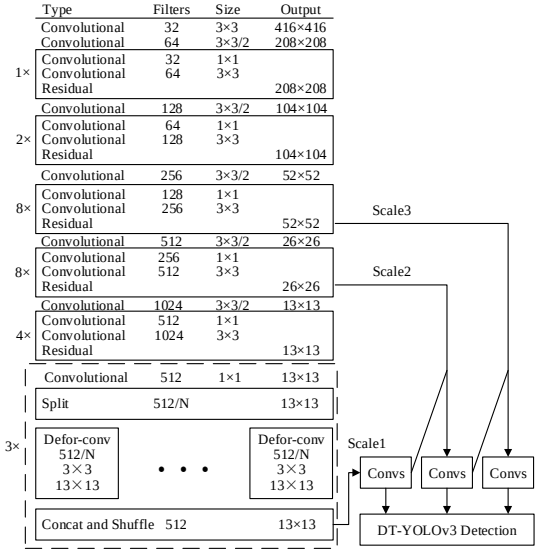


Figure 3. Model structure diagram after group convolution is introduced.

2.3.3. Model optimization. The idea of pruning [12] is to remove the relatively "unimportant" and small contribution weights in the model weight matrix. It includes filter pruning and channel pruning. In filter pruning, the weights with less influence in the convolution kernel are re-moved and in channel pruning, convolution kernels with smaller influence are cut off. We use filter pruning to compress the model in this paper. We apply pruning algorithm on the feature extraction layers of the network, and do not prune feature inter-action layers. The compression ratio is between 0 and 1, the larger the ratio, the more the model is compressed.

3. Experimental Results

3.1. Data Set

The original bridge damage images used in the experiments are provided by the Re-search Institute of Highway Ministry of Transport. The damage includes cracks (20), collapse (11), rust (15), and weeds (18). A total of 64 original images were used. The original images are complex and insufficient. Therefore, we used traditional image processing methods such as adding noise, rotating, shifting and blurring to generate 636 additional images. We also used random erasing method to get 300 images. A total of 1,000 bridge damage images were used in our experiments. We randomly selected 800 images for the model training and 200 images for the model testing.

3.2. Evaluation Method and Experimental Environment

Because bridge damage detection and recognition not only requires accurate prediction of the location information of the damaged object in the image, but also needs to correctly classify the damage type, we used average precision (AP) and mean average precision (mAP) to evaluate our model.

The images include four types of bridge damages: cracks, collapses, rusts, and weeds, so the average accuracy of each type is recorded as AP crack, AP collapse, AP rust, and AP weed. The mAP is the average of AP crack, AP collapse, AP rust, and AP weed. In the evaluation of model detection speed, the detection time of a single image is used.

The hardware environment of this experiment is CPU: Intel (R) Xeon (R) CPU E5-2643; memory size: 64G; GPU: NVIDIA Tesla K40m 12G × 2; memory capacity: 11GB × 2; the operating system is Ubuntu 14.04; GPU computing architecture is Cuda8.0; the deep learning platform uses TensorFlow and Keras framework.

3.3. Evaluation Method and Experimental Environment

We added different numbers of deformable convolution layers to the model and the experimental results are shown in Table 1.

Table 1. Comparison of YOLOv3 and DT-YOLOv3 with different number of deformable convolution layers.

Model	Time /ms	AP crack /%	AP collapse /%	AP rust /%	AP weed /%	mAp /%
YOLOv3	52.7	82.6	79.4	81.2	83.6	81.7
DT-YOLOv3-1	55.1	83.1	80.7	81.6	84.5	82.5
DT-YOLOv3-2	57.9	83.4	81.6	81.7	84.7	82.9
DT-YOLOv3-3	59.2	83.8	82.9	82.5	84.4	83.4
DT-YOLOv3-4	62.3	83.9	82.8	82.4	84.4	83.4
DT-YOLOv3-5	65.4	84.0	82.6	82.6	84.6	83.5

In Table 1, DT-YOLOv3-n means DT-YOLOv3 model with n deformable convolution layers. According to the mAP in Table 1, the model accuracy has been greatly improved after the introduction of three deformable convolution layers. The damage features were well learned by the model. As the number of de-formable convolution layers increased to four and five, although the detection and recognition accuracy of the model has also increased, the increase is extremely small and the speed has been reduced. Therefore, in consideration of the model accuracy and speed, a three-layer deformable convolution was selected and added into the YOLOv3 model.

The experimental results of the DT-YOLOv3 model with different number of transferred layers are shown in Table 2.

Table 2. Comparison of YOLOv3 and DT-YOLOv3 with different number of transferred layers.

Model	Time /ms	AP crack /%	AP collapse /%	Ap rust /%	AP weed /%	mAp /%
YOLOv3	59.2	83.8	82.9	82.5	84.4	83.4
DT-YOLOv3-10	59.2	84.3	83.1	83.0	85.6	84.0
DT-YOLOv3-38	59.2	84.2	83.0	82.8	85.0	83.8
DT-YOLOv3-62	59.2	83.9	82.7	82.7	85.0	83.6
DT-YOLOv3-74	59.2	83.5	82.4	82.3	84.8	83.3

In Table 2, DT-YOLOv3-n means DT-YOLOv3 with n transferred layers. It can be seen from Table 2 that the introduction of transfer learning can effectively improve the performance of the model, and the accuracy of the detection and recognition of weeds increases the most. The loss function values of the model with different transfer layers are shown in Figure 4.

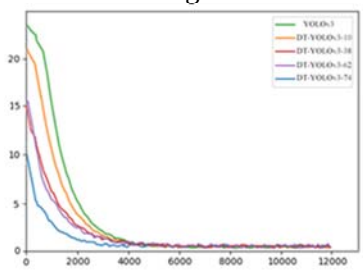


Figure 4. Loss function values of the models with different transfer layers.

Table 2 and Figure 4 show that YOLOv3 convergence speed is the slowest. In DT-YOLOv3, as the number of transfer layers increases, the model convergence speed becomes faster and faster. DT-YOLOv3-74(74 layers are transferred) is the fastest, but in this case the model's detection and recognition rate is the worst. Considering the mAP, we choose to transfer 10 layers in our DT-YOLOv3 model. Part of the bridge damage detection and recognition results are shown in Figure 5.

Group convolution effectively reduces the size of the model, and as the number of convolution groups increases, the model becomes smaller. But the accuracy of the model decreases. When the group number is 2, the accuracy is the maximum.

(a) Rust (b) Collapse (c) Cracks (d) Weed

Figure 5. Bridge damage detection and recognition results.

We trained our DT-YOLOv3 model with group convolution of 2 groups using three different optimizers: Stochastic Gradient Descent (SGD) [13], Nesterov accelerated gradient (NAG) [14] and Adam [15]. The initial learning rate is set to 0.0001. For the NAG algorithm, γ is set to 0.9. In the Adam algorithm, β_1 is set to 0.9, β_2 is set to 0.999, ε is set to 10^{-8} and the batch size is set to 8.

The model trained using the SGD optimizer has the lowest accuracy in bridge damage detection and recognition. In the NAG algorithm, the impact of noise is effectively reduced and he accuracy of the model is greatly improved. Adam's optimizer gets the best performance.

We trained our model with the Adam optimizer, and using pruning compression with different compression ratios. The experimental results are shown in Table 3.

It can be seen that as the compression ratio increases, the model is smaller, but the model's detection and recognition accuracy for bridge damage is also lower. When the compression ratio is 0.1, the model compresses 24.56MB with a little reduce in accuracy. When the compression ratio is 0.2, the model size

reduces a little but the accuracy reduces a lot. This paper uses a model with a compression ratio of 0.1 as the final model to detect and identify bridge damages.

Table 3. Results of DT-YOLOv3 pruning under different compression ratios.

Compression ratio	Model size/MB	Time /ms	AP crack /%	AP collapse /%	AP rust /%	AP weed /%	mAp /%
0.0	239.52	56.9	84.9	83.8	83.8	86.8	84.8
0.1	214.96	47.3	84.6	83.1	83.4	86.6	84.4
0.2	207.47	44.7	81.7	79.9	81.0	82.1	81.2
0.3	189.18	38.8	75.9	75.1	75.7	76.5	75.8
0.4	180.69	35.4	71.5	70.8	70.7	71.5	71.1

4. Conclusion

In this work, we proposed a DT-YOLOv3 model by using deformable convolution and transfer learning. Then we used group convolution and pruning-based strategies to compress the model. The model can extract richer features and it has higher accuracy in bridge damage detection and recognition and is faster compared with YOLOv3. In the future, we will focus on data augmented and few-shot object detection methods. And when we have less training data, our detection model can also achieve better results.

References

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Neural Information Processing Systems, 2012:1097-1105.
[2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Computer Society, 2014: 580-587.
[3] Girshick R. Fast R-CNN[C]. ICCV2015, arXiv:1504.08083.
[4] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelli-gence, 2015, 39(6):1137-1149.
[5] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]. CVPR, 2016, arXiv: 1506.02640v5.
[6] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[C]. Europe-an Conference on Computer Vision. Springer, Cham, 2016: 21-37.
[7] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[J]. 2016: 6517-6525.
[8] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[C]. CVPR, 2018, arXiv: 1804.02767.
[9] LU Y SH, LI Y X, LIU B, LIU H, et al.Hyperspectral data haze monitoring based on deep residual network[J]. ACTC OPTICA SINICA,2017,37(11):314-324.
[10] WANG H B, HAN M, WANG G H, et al. Surfacefeatures extraction in remote sensing images based on architecture - the variant CNN [J]. Acta Geodaetica et Cartographica Sinica, 2019, 13(5):583-596.
[11] HAN D, WANG X J. Pose-varied face recognition based on improved convolutional neural network. Journal of Jilin University (Information Science Edition), 2018, 36(05):376-381.
[12] PENG D L, WANG T X. Pruning algorithm based on GoogLeNet model[J].Control and Decision, 2019, 34(06):1259-1264.
[13] LIU H D, LIU Q, CHEN D H. Adaptive learning-rate on integrated stochastic gradient dcreasing Q-Learning[J]. CHINESE JOURNAL OF COMPUTERS, 2018, 1(12):2019-07-26.
[14] ZHANG Q H, WAN C X, et al. Road object detection algorithm based on convolutional neuralnet work[J]. COMPUTER ENGINEERING AND DESIGN, 2019, 40(07):2052-2058.
[15] YANG G C, YANG C, et al. Modified CNN algorithm is based on dropout and ADAM optimizer[J]. J.Huazhong Univ.Of Sci.& Tech.(Natural Science Edition), 2018, 46(07):122-127.